Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 1929–1940
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Extreme inaccuracies in Gaussian Bayesian networks

Miguel A. Gómez-Villegas[a,*], Paloma Maín[a], Rosario Susi[b]

[a] *Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[b] *Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid, 28040 Madrid, Spain*

Received 9 March 2007
Available online 19 February 2008

Abstract

To evaluate the impact of model inaccuracies over the network's output, after the evidence propagation, in a Gaussian Bayesian network, a sensitivity measure is introduced. This sensitivity measure is the Kullback–Leibler divergence and yields different expressions depending on the type of parameter to be perturbed, i.e. on the inaccurate parameter.

In this work, the behavior of this sensitivity measure is studied when model inaccuracies are extreme, i.e. when extreme perturbations of the parameters can exist. Moreover, the sensitivity measure is evaluated for extreme situations of dependence between the main variables of the network and its behavior with extreme inaccuracies. This analysis is performed to find the effect of extreme uncertainty about the initial parameters of the model in a Gaussian Bayesian network and about extreme values of evidence. These ideas and procedures are illustrated with an example.

AMS 2000 subject classifications: 62F15; 62F35

Keywords: Gaussian Bayesian network; Sensitivity analysis; Kullback–Leibler divergence

Introduction

A Bayesian network is a probabilistic graphical model where a directed acyclic graph (DAG) represents a set of variables with its dependence structure. The nodes are random variables and the edges give dependencies between the variables. A set of conditional distributions of each variable, given its parents, completes the joint description of the variables.

* Corresponding author.
E-mail address: ma.gv@mat.ucm.es (M.A. Gómez-Villegas).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.02.027

Bayesian networks have been studied by some authors, like Pearl [1], Lauritzen [2], Heckerman [3] and Jensen [4], among others.

Depending on the type of variables in the problem, discrete, Gaussian and mixed Bayesian networks can be described.

When a Bayesian network is considered, it is necessary to assign different values to the probabilities of the network. In this step, there can be some uncertainty and then there can exist some inaccurate parameters. Therefore, it is necessary to develop a sensitivity analysis to study how sensitive the network's output is to initial inaccurate parameters.

In the last few years, some methods of sensitivity analysis for Bayesian networks have been developed; for example, for discrete Bayesian networks Laskey [5] presents a sensitivity analysis based on computing the partial derivative of a posterior marginal probability with respect to a given parameter, Coupé and van der Gaag [6] developed an efficient sensitivity analysis based on inference algorithms, and Chan and Darwiche [7] introduced a sensitivity analysis based on a distance measure. For Gaussian Bayesian networks, a sensitivity analysis based on symbolic propagation was developed by Castillo and Kjærulff [8], and on the basis of the Kullback–Leibler divergence, Gómez-Villegas, Maín and Susi [9] proposed a sensitivity measure for performing the sensitivity analysis.

In this paper, we work with the sensitivity measure presented by Gómez-Villegas, Maín and Susi [9] to study its behavior for extreme inaccuracies or perturbations of the parameters that describe the Gaussian Bayesian network. Moreover, we describe the sensitivity measure and its behavior for extreme inaccuracies, when the dependence between the variable of interest X_i and the evidential variable X_e is extreme, that is, when the squared coefficient of correlation between X_i and X_e is given by $\rho_{ie}^2 = 0$ and $\rho_{ie}^2 = 1$, considering different types of relative positioning of these variables in the DAG.

This paper is organized as follows. In Section 1 definitions of Bayesian networks and Gaussian Bayesian networks are introduced and the process of evidence propagation in Gaussian Bayesian networks is reviewed. Also, we present the working example. In Section 2, the sensitivity measure is defined and, depending on the parameter to be perturbed, different expressions of the sensitivity measure are obtained, being able to detect the parameter that perturbs the network's output the most. In Section 3, we study the behavior of the sensitivity measure for extreme perturbations of the initial parameters. In Section 4, we evaluate the sensitivity measure considering extreme situations of dependence between X_i and X_e given by the linear correlation coefficient ρ_{ie} and evaluate those cases with extreme perturbations of the parameters. In Section 5, the sensitivity analysis with the working example is performed for some extreme parameter perturbations considering different positions of the variable of interest X_i and the evidential variable X_e. Finally, in Section 6 some conclusions are presented.

1. Gaussian Bayesian networks and evidence propagation

A Bayesian network and a Gaussian Bayesian network are defined in this section and the process of evidence propagation, one of the most important results in Bayesian networks, is presented. Moreover, these concepts are illustrated with an example.

Definition 1. A Bayesian network is a pair $(\mathcal{G}, \mathcal{P})$ where $\mathcal{G}$ is a DAG, nodes being random variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ and edges probabilistic dependencies between variables, and $\mathcal{P} = \{p(x_1|pa(x_1)), \ldots, p(x_n|pa(x_n))\}$ is a set of conditional probability densities (one for each variable) with $pa(x_i)$ the set of parents of node X_i in $\mathcal{G}$.

The joint probability density of $\mathbf{X}$ is

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|pa(x_i)). \tag{1}$$

Definition 2. A Gaussian Bayesian network is a Bayesian network over $\mathbf{X} = \{X_1, \ldots, X_n\}$ where the joint probability distribution is a multivariate normal distribution $N(\mu, \Sigma)$, with density given by $f(\mathbf{x}) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}$ where μ is the n-dimensional mean vector, Σ the positive definite covariance matrix of dimension $n \times n$ and $|\Sigma|$ the determinant of Σ.

In a Gaussian Bayesian network, the conditional density associated with X_i for $i = 1, \ldots, n$ in Eq. (1) is the univariate normal distribution, with density $f(x_i|pa(x_i)) \sim N(\mu_i + \sum_{j=1}^{i-1} \beta_{ij}(x_j - \mu_j), \nu_i)$ where β_{ij} is the regression coefficient of X_j in the regression of X_i on the parents of X_i with $pa(x_i) \subseteq \{X_1, \ldots, X_{i-1}\}$, and $\nu_i = \Sigma_{ii} - \Sigma_{iPa(x_i)}\Sigma^{-1}_{Pa(xi)}\Sigma'_{iPa(x_i)}$ is the conditional variance of X_i given its parents in the DAG.

It is usual to work with a variable of interest X_i, so the network's output is the information about this variable of interest after the evidence propagation, i.e., the posterior marginal density of X_i.

The evidence propagation is one of the main results associated with Bayesian networks and is the process of updating the probability distribution of the variables of the network introducing the available information about the state of one or more variables, known as evidence variables. Different algorithms have been proposed for propagating the evidence in Bayesian networks.

To perform the evidence propagation in a Gaussian Bayesian network an incremental method is presented, updating one evidential variable at a time [10]. This method is based on computing the conditional probability density of a multivariate normal distribution given the evidential variable X_e. Then, considering the partition $\mathbf{X} = (\mathbf{Y}, E)$, with $\mathbf{Y}$ the set of non-evidential variables, $X_i \in \mathbf{Y}$ being the variable of interest and E the evidential variable, the conditional probability distribution of $\mathbf{Y}$, given the evidence $E = \{X_e = e\}$, is a multivariate normal distribution with parameters

$$\mu^{\mathbf{Y}|E=e} = \mu_{\mathbf{Y}} + \Sigma_{\mathbf{Y}E}\Sigma^{-1}_{EE}(e - \mu_E)$$
$$\Sigma^{\mathbf{Y}|E=e} = \Sigma_{\mathbf{YY}} - \Sigma_{\mathbf{Y}E}\Sigma^{-1}_{EE}\Sigma_{E\mathbf{Y}}.$$

Working with a variable of interest $X_i \in \mathbf{Y}$, after the evidence propagation, we obtain that

$$X_i|E = e \sim N(\mu_i^{\mathbf{Y}|E=e}, \sigma_{ii}^{\mathbf{Y}|E=e}) \equiv N\left(\mu_i + \frac{\sigma_{ie}}{\sigma_{ee}}(e - \mu_e), \sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}}\right) \tag{2}$$

with the parameters before the evidence propagation μ_i and μ_e the means of X_i and X_e, σ_{ii} and σ_{ee} the variances of X_i and X_e, and σ_{ie} the covariance of X_i and X_e.

To illustrate these concepts of Gaussian Bayesian networks and evidence propagation, the following example is introduced.

Example 3. The interest of the problem is in how a machine works. This machine is made up of five elements, the variables of the problem, connected as the DAG presented in Fig. 1. The element of interest in the machine is the last one (X_5).

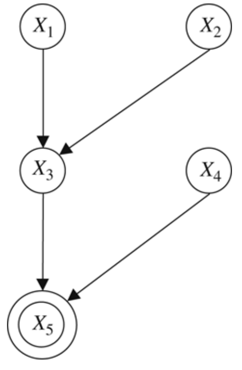

Fig. 1. DAG of the Gaussian Bayesian network.

Let the time for which each element is working be a normal distribution; then $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5\}$ has a multivariate normal distribution given by $\mathbf{X} \sim N(\mu, \Sigma)$, with the conditional parameters described by the DAG

$$\mu = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 5 \end{pmatrix}; \qquad \Sigma = \begin{pmatrix} 3 & 0 & 6 & 0 & 6 \\ 0 & 2 & 2 & 0 & 2 \\ 6 & 2 & 15 & 0 & 15 \\ 0 & 0 & 0 & 2 & 4 \\ 6 & 2 & 15 & 4 & 26 \end{pmatrix}.$$

To obtain Σ, the algorithm presented by Shachter and Kenley [11] is used.

Considering the evidence $E = \{X_2 = 4\}$, after evidence propagation the distribution of the variable of interest is $X_5|X_2 = 4 \sim N(6, 24)$ and the joint probability distribution is a multivariate normal with parameters

$$\mu^{\mathbf{Y}|X_2} = \begin{pmatrix} 2 \\ 4 \\ 4 \\ 6 \end{pmatrix}; \qquad \Sigma^{\mathbf{Y}|X_2} = \begin{pmatrix} 3 & 6 & 0 & 6 \\ 6 & 13 & 0 & 13 \\ 0 & 0 & 2 & 4 \\ 6 & 13 & 4 & 24 \end{pmatrix}.$$

2. Sensitivity analysis in Gaussian Bayesian networks

The sensitivity analysis proposed by Gómez-Villegas, Maín and Susi [9] consists in comparing the network output of two different models: the original model $N(\mu, \Sigma)$ and the perturbed model obtained after adding a *perturbation* $\delta \in \mathbb{R}$ to one inaccurate parameter of the model.

With an evidential variable X_e, whose value is known: $E = \{X_e = e\}$, the evidence propagation is performed in the original model and in the perturbed model, obtaining the network's output as the marginal densities of interest $f(x_i|e)$, for the original model, and $f(x_i|e, \delta)$, for the perturbed model.

Finally, to evaluate the effect of adding a perturbation, the sensitivity measure is used. This measure is the Kullback–Leibler discrepancy [12] used to compare these conditional densities of the variable of interest after the evidence propagation. The Kullback–Leibler divergence for f and f' density functions over the same domain is defined as follows:

$$KL(f(w), f'(w)) = \int_{-\infty}^{+\infty} f(w) \ln \frac{f(w)}{f'(w)} \mathrm{d}w.$$

This measure has been used from the beginning in statistical inference by Jeffreys, Fisher and Lindley. It takes into account the whole behavior of the distributions to be considered. Furthermore, it is very useful when there is no idea about which properties of the variable of interest will be used.

The sensitivity measure is defined as follows.

Definition 4. Let $(\mathcal{G}, \mathcal{P})$ be a Gaussian Bayesian network $N(\mu, \Sigma)$. Let $f(x_i|e)$ be the marginal density of interest after evidence propagation and $f(x_i|e, \delta)$ the same density when the perturbation δ is added to one parameter of the initial model. Then, the sensitivity measure is defined by

$$S^{p_j}(f(x_i|e), f(x_i|e,\delta)) = \int_{-\infty}^{+\infty} f(x_i|e) \ln \frac{f(x_i|e)}{f(x_i|e,\delta)} \mathrm{d}x_i \tag{3}$$

where the subscript p_j is the inaccurate parameter and δ the proposed perturbation, being the new value of the parameter $p_j^\delta = p_j + \delta$.

For small values of the sensitivity measure it is possible to conclude that the Gaussian Bayesian network is robust against the kind of perturbation proposed.

Considering the parameters of the mean vector and the parameters of the covariance matrix as different, and having $\rho_{ie}^2 \in (0, 1)$, the following results are obtained. These results are generally true for conditional distributions in the case of a joint multivariate normal distribution; however their full meaning is reached for Gaussian Bayesian Networks. In fact the conditional distribution, given some evidence, is the output in this kind of representation. Since the main point is to carry out a sensitivity analysis, the effect of perturbations on the conditional distributions has to be studied.

2.1. Mean vector inaccuracy

Depending on the element of μ to be perturbed, the perturbation can affect the mean of the variable of interest $X_i \in \mathbf{Y}$, the mean of the evidential variable $X_e \in E$ or the mean of any other variable $X_j \in \mathbf{Y}$ with $j \neq i$.

Proposition 5. *Considering the perturbation $\delta \in \mathbb{R}$ added to any element of the mean vector μ, and having $\rho_{ie}^2 \in (0, 1)$, the sensitivity measure (3) is as follows:*

(i) *When the perturbation is added to the mean of X_i, then $\mu_i^\delta = \mu_i + \delta$; the density of the variable of interest after the evidence propagation is $X_i|E = e, \delta \sim N(\mu_i^{Y|E=e} + \delta, \sigma_{ii}^{Y|E=e})$, where*

$$S^{\mu_i}(f(x_i|e), f(x_i|e,\delta)) = \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}}.$$

(ii) *If the perturbation is added to the mean of the evidential variable, $\mu_e^\delta = \mu_e + \delta$; the posterior density of the variable of interest after the evidence propagation is $X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e} - \frac{\sigma_{ie}}{\sigma_{ee}}\delta, \sigma_{ii}^{Y|E=e}\right)$, and*

$$S^{\mu_e}(f(x_i|e), f(x_i|e,\delta)) = \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}} \left(\frac{\sigma_{ie}}{\sigma_{ee}}\right)^2.$$

(iii) *The perturbation δ added to the mean of any other non-evidential variable, different from the variable of interest, has no influence over X_i; then $f(x_i|e,\delta) = f(x_i|e)$ and the sensitivity measure is zero.*

When the evidence about X_e is inaccurate, with $e^\delta = e + \delta$, the sensitivity measure obtained is similar to $S^{\mu_e}(f(x_i|e), f(x_i|e,\delta))$. Therefore, this case is studied when the mean of the evidential variable is inaccurate.

2.2. Covariance matrix inaccuracy

When the covariance matrix is perturbed, the structure of the network can change. These changes are in the precision matrix of the perturbed network, i.e., in the inverse covariance matrix considering the perturbation δ.

To guarantee the normality of the network, Σ^δ and $\Sigma^{\mathbf{Y}|E=e,\delta}$ must be positive definite matrices; this restriction, imposed in the following proposition, yields different constraints for the perturbation δ.

Proposition 6. *Adding the perturbation $\delta \in \mathbb{R}$ to the covariance matrix Σ, with $\rho_{ie}^2 \in (0,1)$, the sensitivity measure (3) is as follows:*

(i) *If the perturbation is added to the variance of the variable of interest, then $\sigma_{ii}^\delta = \sigma_{ii} + \delta$ with $\delta > -\sigma_{ii} + \frac{\sigma_{ie}^2}{\sigma_{ee}}$ and after the evidence propagation $X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e}, \sigma_{ii}^{Y|E=e,\delta}\right)$ where $\sigma_{ii}^{Y|E=e,\delta} = (\sigma_{ii} + \delta) - \frac{\sigma_{ie}^2}{\sigma_{ee}}$ and*

$$S^{\sigma_{ii}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[\ln\left(1 + \frac{\delta}{\sigma_{ii}^{Y|E=e}}\right) - \frac{\delta}{\sigma_{ii}^{Y|E=e,\delta}}\right].$$

(ii) *When the perturbation is added to the variance of the evidential variable, with $\sigma_{ee}^\delta = \sigma_{ee} + \delta$ and $\delta > -\sigma_{ee}(1 - \max_{X_j \in \mathbf{Y}} \rho_{je}^2)$ with ρ_{je} the corresponding correlation coefficient, the posterior density of interest is $X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e,\delta}, \sigma_{ii}^{Y|E=e,\delta}\right)$ with $\mu_i^{Y|E=e,\delta} = \mu_i + \frac{\sigma_{ie}^2}{\sigma_{ee}+\delta}(e - \mu_e)$ and $\sigma_{ii}^{Y|E=e,\delta} = \sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}+\delta}$ and the sensitivity measure is given by*

$$S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[\ln\left(\frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}}\right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}}\left(\frac{-\delta}{\sigma_{ee}+\delta}\right)\left(1 + (e-\mu_e)^2\left(\frac{-\delta}{(\sigma_{ee}+\delta)\sigma_{ee}}\right)\right)}{\sigma_{ii}^{Y|E=e,\delta}}\right].$$

(iii) *The perturbation δ added to the variance of any other non-evidential variable $X_j \in \mathbf{Y}$ with $j \neq i$, with $\sigma_{jj}^\delta = \sigma_{jj} + \delta$, has no influence over X_i; then $f(x_i|e,\delta) = f(x_i|e)$ and the sensitivity measure is zero.*

(iv) *When the covariance of the variable of interest X_i and the evidential variable X_e is perturbed, $\sigma_{ie}^\delta = \sigma_{ie} + \delta = \sigma_{ei}^\delta$ and, considering the restriction over δ given as $-\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}} < \delta < -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$, then $X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e,\delta}, \sigma_{ii}^{Y|E=e,\delta}\right)$ with $\mu_i^{Y|E=e,\delta} = \mu_i + \frac{(\sigma_{ie}+\delta)}{\sigma_{ee}}(e - \mu_e)$ and $\sigma_{ii}^{Y|E=e,\delta} = \sigma_{ii} - \frac{(\sigma_{ie}+\delta)^2}{\sigma_{ee}}$. The sensitivity measure* *obtained is*

$$S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[\ln\left(1 - \frac{\delta^2 + 2\sigma_{ie}\delta}{\sigma_{ee}\sigma_{ii}^{Y|E=e}}\right) + \frac{\sigma_{ii}^{Y|E=e} + \left(\frac{\delta}{\sigma_{ee}}(e-\mu_e)\right)^2}{\sigma_{ii}^{Y|E=e,\delta}} - 1\right].$$

(v) *The perturbation δ added to any other covariance, i.e., of X_i and any other non-evidential variable X_j, or of the evidence variable X_e and $X_j \in \mathbf{Y}$ with $j \neq i$, has no influence over the variable of interest; then $f(x_i|e,\delta) = f(x_i|e)$ and the sensitivity measure is zero.*

The proof and more details about the sensitivity analysis proposed with an example can be seen in Gómez-Villegas, Maín and Susi [9].

3. Extreme behavior of the sensitivity measure

To know the effect of extreme uncertainty about the initial parameters of the network, we study the behavior of the sensitivity measure for extreme perturbations with the limit of the sensitivity measure. In this case, the squared correlation coefficient of X_i and X_e considered is $\rho_{ie}^2 \in (0,1)$.

Proposition 7. *When the perturbation added to the mean vector is extreme and $\rho_{ie}^2 \in (0,1)$, the sensitivity measure is extreme too and it is:*

(i) (a) $\lim_{\delta\to\pm\infty} S^{\mu_i}(f(x_i|e), f(x_i|e,\delta)) = +\infty$,
(b) $\lim_{\delta\to 0} S^{\mu_i}(f(x_i|e), f(x_i|e,\delta)) = 0$;
(ii) (a) $\lim_{\delta\to\pm\infty} S^{\mu_e}(f(x_i|e), f(x_i|e,\delta)) = +\infty$,
(b) $\lim_{\delta\to 0} S^{\mu_e}(f(x_i|e), f(x_i|e,\delta)) = 0$.

Proof. If follows directly. □

Therefore the limit for an extreme value of the evidence e corresponds to the case (ii).

Proposition 8. *When the extreme perturbation is added to the elements of the covariance matrix and $\rho_{ie}^2 \in (0,1)$, the sensitivity measure is as follows:*

(i) (a) $\lim_{\delta\to+\infty} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e,\delta)) = +\infty$ *but* $S^{\sigma_{ii}}(f(x_i|e), f(x_i|e,\delta)) = o(\delta)$,
(b) $\lim_{\delta\to M_{ii}} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e,\delta)) = +\infty$ *with* $M_{ii} = -\sigma_{ii} + \frac{\sigma_{ie}^2}{\sigma_{ee}} = -\sigma_{ii}\left(1-\rho_{ie}^2\right)$,
(c) $\lim_{\delta\to 0} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e,\delta)) = 0$;
(ii) (a) $\lim_{\delta\to+\infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[-\ln\left(1-\rho_{ie}^2\right) - \rho_{ie}^2\left(1 - \frac{(e-\mu_e)^2}{\sigma_{ee}}\right)\right]$,
(b)

$$\lim_{\delta\to M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[\ln\left(\frac{M_{ee}^* - \rho_{ie}^2}{M_{ee}^*(1-\rho_{ie}^2)}\right) + \frac{\rho_{ie}^2(1-M_{ee}^*)}{M_{ee}^* - \rho_{ie}^2}\left(1 + \frac{(e-\mu_e)^2}{\sigma_{ee}}\left(\frac{1-M_{ee}^*}{M_{ee}^*}\right)\right)\right]$$

where $M_{ee} = -\sigma_{ee}(1-M_{ee}^*)$ *and* $M_{ee}^* = \max_{X_j\in\mathbf{Y}} \rho_{je}^2$;
if $M_{ee}^* = \rho_{ie}^2$ *then* $\lim_{\delta\to M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = +\infty$,
(c) $\lim_{\delta\to 0} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = 0$;
(iii) (a) $\lim_{\delta\to M_{ie}^1} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta)) = +\infty$ *with* $M_{ie}^1 = -\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}}$,

(b) $\lim_{\delta \to M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta)) = +\infty$ *with* $M_{ie}^2 = -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$,

(c) $\lim_{\delta \to 0} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta)) = 0$.

Proof. (i) (a) and (c) follow directly.

(b) When $\sigma_{ii}^{\delta} = \sigma_{ii} + \delta$ the new variance of X_i is $\sigma_{ii}^{\mathbf{Y}|E=e,\delta} = \sigma_{ii}^{\mathbf{Y}|E=e} + \delta$.

Considering $\sigma_{ii}^{\mathbf{Y}|E=e,\delta} > 0$ then $\delta > -\sigma_{ii}^{\mathbf{Y}|E=e}$.

Defining $M_{ii} = -\sigma_{ii}^{\mathbf{Y}|E=e}$ and with $x = \sigma_{ii}^{\mathbf{Y}|E=e} + \delta$, we have

$$\lim_{\delta \to M_{ii}} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e,\delta)) = \lim_{x \to 0} \frac{1}{2}\left[\ln x - \ln \sigma_{ii}^{\mathbf{Y}|E=e} - \frac{x - \sigma_{ii}^{\mathbf{Y}|E=e}}{x}\right] = +\infty.$$

(ii) (a) $\lim_{\delta \to +\infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[\ln\left(\frac{\sigma_{ii}}{\sigma_{ii}^{Y|E=e}}\right) + \frac{-\frac{\sigma_{ie}^2}{\sigma_{ee}}\left(1 - \frac{(e-\mu_e)^2}{\sigma_{ee}}\right)}{\sigma_{ii}}\right]$ with $\sigma_{ii}^{Y|E=e} = \sigma_{ii}(1 - \rho_{ie}^2)$ and $\rho_{ie}^2 = \frac{\sigma_{ie}^2}{\sigma_{ii}\sigma_{ee}}$; then

$$\lim_{\delta \to +\infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta)) = \frac{1}{2}\left[-\ln\left(1 - \rho_{ie}^2\right) - \rho_{ie}^2\left(1 - \frac{(e-\mu_e)^2}{\sigma_{ee}}\right)\right]$$

(b) When $\sigma_{ee}^{\delta} = \sigma_{ee} + \delta$, the new conditional variance for all non-evidential variables is $\sigma_{jj}^{Y|E=e,\delta} = \sigma_{jj} - \frac{\sigma_{je}^2}{\sigma_{ee}+\delta}$. If it is imposed that $\sigma_{jj}^{Y|E=e,\delta} > 0$ for all $X_j \in \mathbf{Y}$ then δ must satisfy following condition: $\delta > -\sigma_{ee}(1 - \max_{X_j \in \mathbf{Y}} \rho_{je}^2)$.

Defining $M_{ee}^* = \max_{X_j \in \mathbf{Y}} \rho_{je}^2$ and $M_{ee} = -\sigma_{ee}(1 - M_{ee}^*)$,

$$\lim_{\delta \to M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta))$$
$$= \lim_{\delta \to M_{ee}} \frac{1}{2}\left[\ln\left(\frac{\sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}+\delta}}{\sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}}}\right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}}\left(\frac{-\delta}{\sigma_{ee}+\delta}\right)\left(1 + (e-\mu_e)^2\left(\frac{-\delta}{(\sigma_{ee}+\delta)\sigma_{ee}}\right)\right)}{\sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}+\delta}}\right]$$
$$= \frac{1}{2}\left[\ln\left(\frac{\sigma_{ii}\sigma_{ee}M_{ee}^* - \sigma_{ie}^2}{M_{ee}^*(\sigma_{ii}\sigma_{ee} - \sigma_{ie}^2)}\right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}}\left(\frac{1-M_{ee}^*}{M_{ee}^*}\right)\left(1 + \frac{(e-\mu_e)^2}{\sigma_{ee}}\left(\frac{1-M_{ee}^*}{M_{ee}^*}\right)\right)}{\left(\frac{\sigma_{ii}}{M_{ee}^*}\right)M_{ee}^* - \rho_{ie}^2}\right]$$
$$= \frac{1}{2}\left[\ln\left(\frac{M_{ee}^* - \rho_{ie}^2}{M_{ee}^*(1 - \rho_{ie}^2)}\right) + \frac{\rho_{ie}^2(1 - M_{ee}^*)}{M_{ee}^* - \rho_{ie}^2}\left(1 + \frac{(e-\mu_e)^2}{\sigma_{ee}}\left(\frac{1 - M_{ee}^*}{M_{ee}^*}\right)\right)\right].$$

If $M_{ee}^* = \rho_{ie}^2 \neq 0$ then $M_{ee} = -\sigma_{ee}(1 - \rho_{ie}^2) = -\sigma_{ee} + \frac{\sigma_{ie}^2}{\sigma_{ii}}$; it follows that

$$\lim_{\delta \to M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e,\delta))$$
$$= \lim_{\delta \to M_{ee}} \frac{1}{2}\left[\ln\left(\frac{M_{ee}^* - \rho_{ie}^2}{M_{ee}^*(1 - \rho_{ie}^2)}\right) + \frac{\rho_{ie}^2(1 - M_{ee}^*)}{M_{ee}^* - \rho_{ie}^2} \times \left(1 + \frac{(e-\mu_e)^2}{\sigma_{ee}}\left(\frac{1 - M_{ee}^*}{M_{ee}^*}\right)\right)\right] = +\infty$$

because for $K, M > 0$

$$\lim_{\substack{x\to 0\\ x>0}}\left[\ln\frac{x}{M}+\frac{K}{x}\right]=\lim_{x\to 0}\frac{1}{x}\left[x\ln\frac{x}{M}+K\right]=\lim_{x\to 0}\frac{1}{x}\left[\frac{\ln\frac{x}{M}}{\frac{1}{x}}+K\right]=+\infty.$$

(c) It follows directly for the case $M_{ee}^{*}=1$.

(iii) (a) For $\sigma_{ie}^{\delta}=\sigma_{ie}+\delta$, the new conditional variance is $\sigma_{ii}^{Y|E=e,\delta}=\sigma_{ii}-\frac{(\sigma_{ie}+\delta)^2}{\sigma_{ee}}$; then, if $\sigma_{ii}^{Y|E=e,\delta}>0$ is imposed, δ must satisfy the following condition: $-\sigma_{ie}-\sqrt{\sigma_{ii}\sigma_{ee}}<\delta<-\sigma_{ie}+\sqrt{\sigma_{ii}\sigma_{ee}}$.

First, defining $M_{ie}^2=-\sigma_{ie}+\sqrt{\sigma_{ii}\sigma_{ee}}$, it is possible to calculate $\lim_{\delta\to M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta))$. But $\delta\to M_{ie}^2$ is equivalent to $\left(\delta^2+2\sigma_{ie}\delta\right)\to\sigma_{ee}\sigma_{ii}^{Y|E=e}$ and given that

$$S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta))=\frac{1}{2}\left[\ln\left(\frac{\sigma_{ee}\sigma_{ii}^{Y|E=e}-(\delta^2+2\sigma_{ie}\delta)}{\sigma_{ee}\sigma_{ii}^{Y|E=e}}\right)+\frac{\sigma_{ee}\sigma_{ii}^{Y|E=e}+\left(\frac{\delta}{\sigma_{ee}}(e-\mu_e)\right)^2}{\sigma_{ee}\sigma_{ii}^{Y|E=e}-(\delta^2+2\sigma_{ie}\delta)}-1\right]$$

and using $\lim_{x\to 0}\left[\ln x+\frac{k}{x}\right]=+\infty$ for every k, then

$$\lim_{\delta\to M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e,\delta))=+\infty.$$

(b) Analogously to (a).

(c) It follows directly. □

The results obtained are the expected ones, except when the extreme perturbation is added to the evidential variance having a finite limit. This is because the state of the evidential variable is known and the variance of this variable has a limited effect on the variable of interest. In this case, the posterior density of interest with the perturbation in the model $f(x_i|e,\delta)$ is not so different to the posterior density of interest without the perturbation $f(x_i|e)$. Therefore, although an extreme perturbation added to the evidential variance can exist, the sensitivity measure tends to a finite value.

4. Extreme dependence between the variable of interest and the evidential variable

In this section we evaluate some particular cases of the sensitivity measure and its extreme behavior, depending on the relations between the variable of interest X_i and the evidential variable X_e.

Then, considering extreme values for the squared linear correlation coefficient, for example $\rho_{ie}^2=0$, X_i and X_e are independent and the connection between these variables in the DAG must be a converging connection (only this connection considers independence between the variables). On the other hand, with $\rho_{ie}^2=1$, the connection in the DAG between X_i and X_e can be a serial or diverging connection, with a linear dependence relation between X_i and X_e.

If X_i and X_e are independent, when the evidence propagation is done, the information about X_e does not affect the variable of interest; then after the evidence propagation $\mu_i^{Y|E=e}=\mu_i$ and $\sigma_{ii}^{Y|E=e}=\sigma_{ii}$.

In this case, only when the perturbation is added to the parameters of X_i is the sensitivity measure not zero. Moreover, when the covariance between X_i and X_e is perturbed, the relation between those variables changes to $\sigma_{ie}=\delta$ and therefore the structure of the network changes. In those cases the sensitivity measure is obtained considering $\sigma_{ie}=0$ in Propositions 5 and 6. Any other perturbation of the parameters of the network does not disturb the results about X_i.

The other extreme case for the squared correlation coefficient is $\rho_{ie}^2 = 1$. In this case, any perturbation associated with the initial parameters of X_i or X_e changes significantly the results concerning X_i; then the sensitivity measure computed in all cases is infinity. Therefore, for a linear relation between the evidential variable and the variable of interest, and then a maximum value of ρ_{ie}^2, the sensitivity measure is also extreme.

When X_i and X_e are independent, only perturbations associated with the parameters of X_i can disturb the network's output, and if the perturbations are extreme, the sensitivity measure is also extreme. However, for $\rho_{ie}^2 = 1$, any perturbation added to the parameters of X_i or X_e greatly affects the network's output, the sensitivity measure being infinity.

5. Working example

Example 9. Consider the Gaussian Bayesian network in Example 3. Experts disagree with the parameters assigned to the joint probability distribution and they want to quantify the effect of inaccuracy when some extreme perturbations of the parameters are proposed.

Let the mean and the variance of the variable of interest X_5 be $\mu_5^{\delta_5} = -20 = \mu_5 + \delta_5$ (with $\delta_5 = -25$) and $\sigma_{55}^{\delta_{55}} = 3$ (with $\delta_{55} = -23$). Let the mean and the variance of the evidential variable X_2 be given by $\mu_2^{\delta_2} = 30 = \mu_2 + \delta_2$ (with $\delta_2 = 27$) and $\sigma_{22}^{\delta_{22}} = 0.27$ (with $\delta_{22} = -1.73$). Finally, let $\sigma_{52}^{\delta_{52}} = 3$ with $\delta_{52} = 1$.

The sensitivity measure yielded for these inaccuracy parameters is

$$S^{\mu_5}(f(x_5|X_2 = 4), f(x_5|X_2 = 4, \delta_5)) = 13.02$$
$$S^{\sigma_{55}}(f(x_5|X_2 = 4), f(x_5|X_2 = 4, \delta_{55})) = 9.91$$
$$S^{\mu_2}(f(x_5|X_2 = 4), f(x_5|X_2 = 4, \delta_2)) = 15.19$$
$$S^{\sigma_{22}}(f(x_5|X_2 = 4), f(x_5|X_2 = 4, \delta_{22})) = 2.03$$

(where 2.1213 is the limit of the sensitivity measure when δ_{22} tends to M_{ee});

$$S^{\sigma_{52}}(f(x_5|X_2 = 4), f(x_5|X_2 = 4, \delta_{52})) = 0.009.$$

It can be pointed out that although there can exist more perturbed parameters, those inaccuracies do not disturb the network's output.

To show the behavior of the measure, we present the sensitivity measures as a function of the perturbation δ in Fig. 2.

In this case, it can be noted that the sensitivity measure of the mean of X_i is the same as the sensitivity measure of the mean of X_e, because of the initial parameters that have been chosen.

As can be seen in the example, the sensitivity measure grows when the perturbation is large. When the evidential variance is perturbed it is necessary to compute the limit of the measure to know whether the perturbation proposed is also large.

Example 10. Consider the Gaussian Bayesian network given in Fig. 1, the variable of interest being X_1, with the same evidential variable X_2. The variables X_1 and X_2 are in a converging connection in the graph, being independent variables. In this case, if any parameter different to μ_1, σ_{11} or σ_{12} is perturbed, this inaccuracy does not affect the posterior density of X_1; then the sensitivity measure is zero. However, if μ_1, σ_{11} or σ_{12} are perturbed, the sensitivity measure is different to zero.

Therefore, in this case it is important to be very careful in assigning the parameters of the variable of interest X_1.

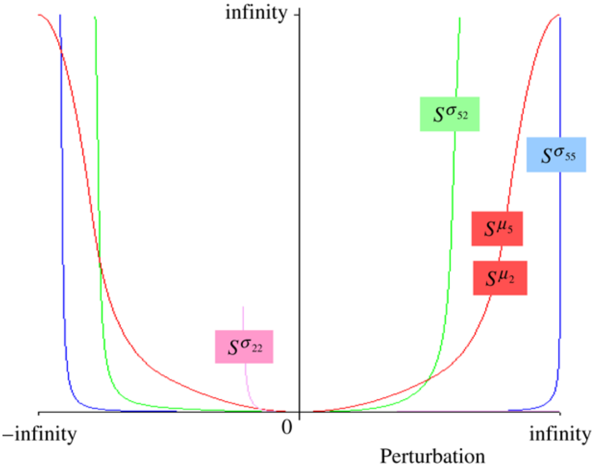

Fig. 2. Sensitivity measures obtained in the example for any perturbation.

6. Conclusions

In this paper the behavior of the sensitivity measure introduced by Gómez-Villegas, Maín and Susi [9] has been studied. This measure is useful for evaluating the impact of parameter inaccuracies in a Gaussian Bayesian network over the density of interest after the evidence propagation, when the inaccuracies are extreme. With this analysis it is possible to prove that this is a well-defined measure for developing a sensitivity analysis in a Gaussian Bayesian network even if the proposed perturbations are extreme.

The results obtained are the expected ones. When the evidential variance is inaccurate with a large perturbation associated with this value, there exists a finite value as the limit of the sensitivity measure. This is because the evidence about this variable explains the behavior of the variable of interest regardless of its inaccurate variance.

Moreover, in all possible cases for the sensitivity measure, if the perturbation added to a parameter is very small, tending to zero, the sensitivity measure is also zero.

Furthermore, it is possible to evaluate the sensitivity measure in some particular cases depending on the dependence relation between the variable of interest X_i and the evidential variable X_e. In that way, $\rho_{ie}^2 \in (0, 1)$ and also extreme values of the linear squared correlation coefficient ($\rho_{ie}^2 = 0$ and $\rho_{ie}^2 = 1$) are considered, having different types of connections in the DAG.

If there is no relation between X_i and X_e, the connection in the graph is a converging connection. Therefore the evidence and any perturbations added to parameters of the evidential variable X_e do not affect the information about the variable of interest X_i. In this case, the sensitivity measure is different to zero only when the parameters of X_i are inaccurate.

For a linear relation between X_i and X_e, with $\rho_{ie}^2 = 1$, having a serial or diverging connection in the DAG, the sensitivity measure is infinity because any perturbation about X_i or X_e produces an important effect over the network's output. Therefore, if X_e is so related to X_i, then any perturbation added to the parameters that describe these variables makes the sensitivity measure extreme.

Acknowledgments

This research was supported by the MEC from Spain, Grant MTM2005-05462, and the Universidad Complutense-Comunidad de Madrid, Grant UCM2005-910395.

References

[1] J. Pearl, Probabilistic reasoning in intelligent systems, in: Networks of Plausible Inference, Morgan Kaufmann, Palo Alto, 1988.
[2] S.L. Lauritzen, Graphical Models, Clarendon Press, Oxford, 1996.
[3] D. Heckerman, A tutorial on learning with Bayesian networks, in: M.I. Jordan (Ed.), Learning in Graphical Models, MIT Press, Cambridge, MA, 1998.
[4] F.V. Jensen, Bayesian Networks and Decision Graphs, Springer, New York, 2001.
[5] K.B. Laskey, Sensitivity analysis for probability assessments in Bayesian networks, IEEE Transactions on Systems, Man, and Cybernetics 25 (6) (1995) 901–909.
[6] V.M.H. Coupé, L.C. van der Gaag, Properties of sensitivity analysis of Bayesian belief networks, Annals of Mathematics and Artificial Intelligence 36 (2002) 323–356.
[7] H. Chan, A. Darwiche, A distance measure for bounding probabilistic belief change, International Journal of Approximate Reasoning 38 (2) (2005) 149–174.
[8] E. Castillo, U. Kjærulff, Sensitivity analysis in Gaussian Bayesian networks using a symbolic–numerical technique, Reliability Engineering and System Safety 79 (2003) 139–148.
[9] M.A. Gómez-Villegas, P. Maín, R. Susi, Sensitivity analysis in Gaussian Bayesian networks using a divergence measure, Communications in Statistics–Theory and Methods 36 (3) (2007) 523–539.
[10] E. Castillo, J.M. Gutiérrez, A.S. Hadi, Expert Systems and Probabilistic Network Models, Springer-Verlag, New York, 1997.
[11] R. Shachter, C. Kenley, Gaussian influence diagrams, Management Science 35 (5) (1989) 527–550.
[12] S. Kullback, R.A. Leibler, On information and sufficiency, Annals of Mathematical Statistics 22 (1951) 79–86.